

Natural Language Processing Theory, Applications and Difficulties

Mrs. Anjali Gharat¹, Mrs. Helina Tandel², Mr. Ketan Bagade³

¹Faculty of Computer Engineering,

²Faculty of Electronics & Telecommunications,

³Faculty of Information Technology,

^{1,2,3}Vidyalankar Polytechnic, Wadala, Mumbai, India

ABSTRACT

The promise of a powerful computing device to help people in productivity as well as in recreation can only be realized with proper human-machine communication. Automatic recognition and understanding of spoken language is the first step toward natural human-machine interaction. Research in this field has produced remarkable results, leading to many exciting expectations and new challenges. This field is known as Natural language Processing. In this paper the natural language generation and Natural language understanding is discussed. Difficulties in NLU, applications and comparison with structured programming language are also discussed here.

KEYWORDS: Natural Language Processing; Natural Language Understanding, Natural Language Generation

How to cite this paper: Mrs. Anjali Gharat | Mrs. Helina Tandel | Mr. Ketan Bagade "Natural Language Processing Theory, Applications and Difficulties" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-3 | Issue-6, October 2019, pp.501-503, URL: <https://www.ijtsrd.com/papers/ijtsrd28092.pdf>



Copyright © 2019 by author(s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (<http://creativecommons.org/licenses/by/4.0>)



I. INTRODUCTION

Artificial intelligence (AI) is changing the way we look at the world. AI "robots" are everywhere. From our phones to devices like Amazon's Alexa. The field of study that focuses on the interactions between human language and computers is called Natural Language Processing. Natural Language Processing (NLP) refers to AI method of communicating with intelligent systems using a natural language such as English. NLP is a way for computers to analyze, understand, and derive meaning from human language in a smart and useful way. By utilizing NLP, developers can organize and structure knowledge to perform tasks such as automatic summarization, translation, named entity recognition, relationship extraction, sentiment analysis, speech recognition, and topic segmentation.

II. Related Works

In [1] an interactive approach to ease the construction, review, and revision of Natural Language Processing models for extractive summarization is presented. An interactive tool to be used by clinicians to identify text for inclusion in brief patient summaries ("sign-out notes") from full-text patient notes is presented. Physicians and nurses use sign-out notes to provide concise patient summaries and are used to facilitate transitions in care between providers. These sign-out notes summarize key observations and interpretations from a patient's medical record. Although

some health care organization has adopted electronic sign-out systems, these tools do not automate the process. Even when prepared on a computer, sign-out notes must be prepared manually, multiple times throughout each patient's hospitalization, making sign-out prone to errors and omissions. Different teams of physicians have different needs and practices regarding the contents of a useful sign-out, and need customized training examples to build individualized models. An intelligent tool for clinicians to build NLP models for identifying text (within patient notes) for inclusion in a sign-out note is proposed here.

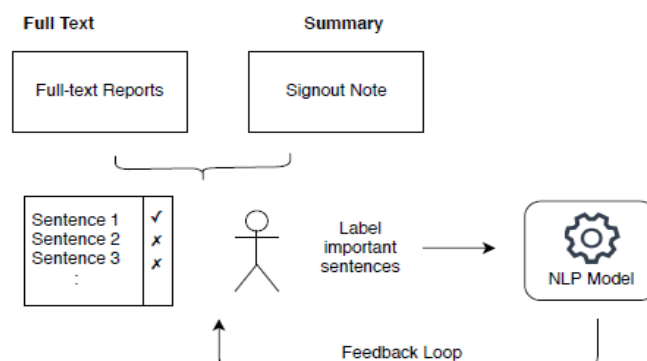


Figure1: Building NLP models interactively, for identifying text for inclusion in sign-out notes.

In [2], a Text-to speech synthesizer is developed that converts text into spoken word, by analyzing and processing it using Natural Language Processing (NLP) and then using Digital Signal Processing (DSP) technology to convert this processed text into synthesized speech representation of the text.

A TTS system (or "voice generated-engine") is composed of one interface as a front-end and a back-end. The first interface converts raw text containing alpha-numeric symbols like numbers and abbreviations in terms of speech into the equivalent of out words. Here analysis of text includes various features such as recognition of text unit, normalization of text unit pattern, pre-processing, etc. This process of transcriptions is known as text-to-phoneme (TTP) or grapheme-to-phoneme (GTP) conversion. These two conversions are together known as symbolic linguistic representation which is the desired output of TTS.

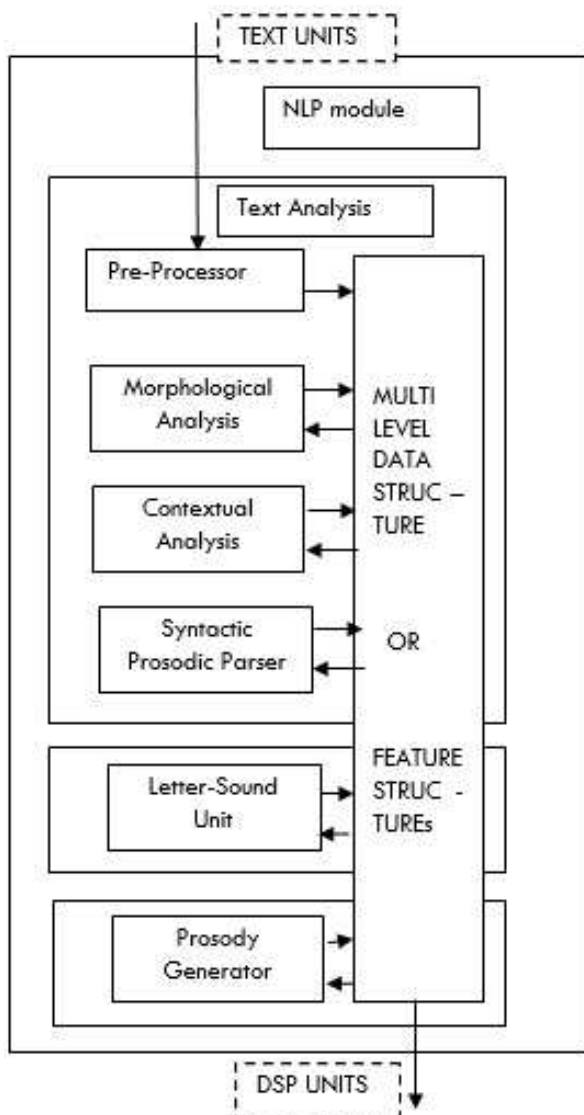


Figure 2: TTS synthesis model

In [3] proposed approach is to develop a mobile based app about agriculture process to get the optimal Solution from owl files based on text processing using SPARQL query from NLP language

III. Natural Language Processing

In this topic Natural Language Processing (NLP) fundamentals are discussed.

What is NLP?

The field of NLP involves making computers to perform useful tasks with the natural languages humans use. The input and output of an NLP system can be Speech Written Text.

Components of NLP:

Components of NLP are

Natural Language Understanding (NLU) and Natural Language Generation (NLG).

A. Natural Language Understanding (NLU)

Understanding involves the following tasks – Mapping the given input in natural language into useful representation, Analyzing different aspects of the language.

B. Natural Language Generation (NLG)

It is the process of producing meaningful phrases and sentences in the form of natural language from some internal representation.

Text planning – It includes retrieving the relevant content from knowledge base.

Sentence planning – It includes choosing required words, forming meaningful phrases, setting tone of the sentence.

Text Realization – It is mapping sentence plan into sentence structure.

The NLU is harder than NLG.

C. NLP Terminology

Phonology – It is study of organizing sound systematically.

Morphology – It is a study of construction of words from primitive meaningful units.

Morpheme – It is primitive unit of meaning in a language.

Syntax – It refers to arranging words to make a sentence. It also involves determining the structural role of words in the sentence and in phrases.

Semantics – It is concerned with the meaning of words and how to combine words into meaningful phrases and sentences.

D. Steps in NLP

There are general five steps

1. Lexical Analysis –

It involves identifying and analyzing the structure of words. Lexicon of a language means the collection of words and phrases in a language. Lexical analysis is dividing the whole chunk of text into paragraphs, sentences, and words.

2. Syntactic Analysis (Parsing) –

It involves analysis of words in the sentence for grammar and arranging words in a manner that shows the relationship among the words. The sentence such as "The school goes to boy" is rejected by English syntactic analyzer.

3. Semantic Analysis –

It draws the exact meaning or the dictionary meaning from the text. The text is checked for meaningfulness. It is done by mapping syntactic structures and objects in the task domain. The semantic analyzer disregards sentence such as "hot ice-cream".

4. Discourse Integration –

the meaning of any sentence depends upon the meaning of the sentence just before it. In addition, it also brings about the meaning of immediately succeeding sentence.

5. Pragmatic Analysis –

during this, what was said is re-interpreted on what it actually meant. It involves deriving those aspects of language which require real world knowledge.

IV. Applications

1. Automatic summarization

Information overload is a real problem when we need to access a specific, important piece of information from a huge knowledge base. Automatic summarization is relevant not only for summarizing the meaning of documents and information, but also for understanding the emotional meanings inside the information, such as in collecting data from social media. Automatic summarization is especially relevant when used to provide an overview of a news item or blog posts, while avoiding redundancy from multiple sources and maximizing the diversity of content obtained.

2. Machine Translation

As the **amount of information available online is growing**, the need to access it becomes increasingly important and the value of natural language processing applications becomes clear. Machine translation helps us conquer language barriers.. The challenge with **machine translation technologies is not in translating words, but in understanding the meaning of sentences** to provide a true translation

3. Text classification

Text classification makes it possible to assign predefined categories to a document and organizes it to help you find the information you need or simplify some activities. For example, an application of text categorization is spam filtering in email.

4. Question Answering

As speech-understanding technology and voice-input applications improve, **the need for NLP will only increase**. Question-Answering (QA) is becoming more and more popular. **It may be used as a text-only interface or as a spoken dialog system**

V. Comparison with structured programming Languages:

In the world of computing, information plays an important role in our lives. One of the major sources of information is

database. Database and Database technology are having major impact on the growing use of computers. Almost all IT applications are storing and retrieving the information or data from the database. Database Management Systems (DBMS) have been widely used for storing and retrieving data. For storing and retrieving the information from database requires the knowledge of database language like SQL. However, everyone may not be able to write the SQL query. NLP solves this problem.

VI. Conclusion

NLP is building systems that can understand language. It is a subset of Artificial Intelligence.

The benefits of natural language processing are innumerable. Natural language processing can be used by companies to improve the efficiency of documentation processes, improve the accuracy of documentation, and identify the most important information from large databases.

As we have seen, NLP provides a wide set of techniques and tools which can be applied in all the areas of life. By learning them and using them in our everyday interactions, our life quality would highly improve

VII. References

- [1] Gaurav Trivedi, Robert Handzel, Shyam Visweswaran, "An Interactive NLP Tool for Signout Note Preparation" 2018 IEEE International Conference on Healthcare Informatics.
- [2] Partha Mukherjee, Soumen Santra, Ananya Paul, "Development of GUI for Text-to-Speech Recognition using Natural Language Processing "2018 IEEE International conference"
- [3] Dr. N. Srinivasan, Anandaraj Selvaraj, " Mobile Based Data Retrieval using RDF and NLP in an Efficient Approach", 2017 Third International Conference on Science Technology Engineering & Management (ICONSTEM)
- [4] G. Trivedi, P. Pham, W. W. Chapman, R. Hwa, J. Wiebe, and H. Hochheiser, "NLPReViz: an interactive tool for natural language processing on clinical text," Journal of the American Medical Informatics Association, 2017
- [5] S. Patil, M. Phonde, S. Prajapati, S. Rane, A. Lahane, "Multilingual Speech and Text Recognition and Translation using Image", International Journal of Engineering Research & Technology, vol. 5(4), pp. 85-87, 2016